Technische Universität Ilmenau
Institut für Mathematik

Preprint No. M 10/02

Universal confidence sets for the mode of a regression function

Sinotina, Tatiana; Vogel, Silvia

2010

Impressum:
Hrsg.: Leiter des Instituts für Mathematik
Weimarer Straße 25
98693 Ilmenau
Tel.: +49 3677 69 3621
Fax: +49 3677 69 3270
http://www.tu-ilmenau.de/ifm/

ISSN xxxx-xxxx

ilmedia

Universal confidence sets for the mode of a regression function

Tatiana Sinotina, Silvia Vogel

Abstract. We consider the problem of mode estimation of the regression model with random (stochastic) design. Confidence sets for the modes can be derived as suitable neighborhoods of maximum point of a regression estimator. For each sample size n the neighborhoods are chosen in such a way that they cover the true modes at least with a prescribed probability. The approach relies on concentration-on-measure inequalities for the regression estimators. The aim of the talk is to derive appropriate assertions for the famous regression estimators and to show how they can be used for the determination of universal confidence sets.

MSC classification: 62G08, 90C15 MSC classification: 62G08, 90C15

Key Words: Nonparametric regression, Stochastic programming

1. Introduction. Let (X,Y) be a two-dimensional random vector with unknown continuous bivariate distribution such that Y is integrable. We consider the *random design regression model*:

$$m(x) := E(Y \mid X = x),\ x \in \mathrm{R}. \tag{1.1}$$

X is assumed to have a density g. The aim of this paper is to estimate the location of the maximum of the unknown regression function. The typical procedure in this case consists of three stages. Firstly, the unknown regression model should be approximated. Then this estimation is used to solve the maximization problem and, thus, to obtain the approximate location of maximum. In the end we usually have to answer the question of how the resulting value is close to the "true".

There are several ways to estimate the regression model. In this paper we consider one of the classical estimators. It was independently introduced in the middle of the last century by Nadaraya and Watson:

$$\hat{m}(x) := \frac{\frac{1}{nh_n}\sum_{i=1}^{n} Y_i K(\frac{x-X_i}{h_n})}{\frac{1}{nh_n}\sum_{i=1}^{n} K(\frac{x-X_i}{h_n})} = \frac{\hat{r}(x)}{\hat{g}(x)},\ x \in \chi. \tag{1.2}$$

In this formula K is a kernel function, h_n is a bandwidth and $\chi \subset \mathrm{R}$ is some compact set.

Assume that the maximum of m is situated at some unknown point $x_{max} \in \chi$. Using the Nadaraya-Watson estimator we estimate the value of $m(x_{max})$ by

$$\hat{m}(\hat{x}_{max}) = \max_{x} \hat{m}(x),\ x \in \chi.$$

The idea of estimating location and size of a maximum of a nonparametric curve is closely related to the problem of estimation the mode of a density. The first work in this direction belongs to Parzen. In 1962 he showed under some assumptions on K that $\hat{g}$ is an asymptotically unbiased and consistent estimator for the density g whenever $h \equiv h_n \to 0, nh_n \to \infty$ and x is a continuity point of g. With some additional assumptions on g and h_n an asymptotic normality result can also be obtained. In 1964 Nadaraya established similar results for $\hat{m}$. Since then the kernel estimation was widely used to solve different problems. One of them is the problem of mode estimation - both for the density and the regression model. After this such scientists as E. Patzen, A. Tsybakov, W. Härdle, K.Ziegler etc. touched this subject in their publications.

In this paper we make no attempt to improve any existent regression estimator. We use some common model. On its base we will construct a confidence set for the mode value using uniform concentration-of-measure results and some additional assumptions about the true model. This means that for each sample size n we will derive a random set, which covers the estimated parameter with the prescribed probability. These random sets are called (strong) *universal confidence sets*. The idea of this method was firstly proposed by G. Pflug [Pf04] and was further developed by S. Vogel in [Vog07] and [Vog08].

This paper is organized as follows. In the second section we introduce universal confidence sets and explain the method in general. In section 3 we investigate the Nadaraya-Watson estimator with respect to the convergence properties needed for the derivation of the universal confidence sets. We condine to a single-valued solution set. In Section 4 we extend our results to the case that the function has multiple peaks. And in the end of the paper we discuss the problems of choice of the optimal bandwidth.

2. Universal confidence sets (UCS) In this section we set the main definitions and theorems that are necessary for constructing UCS (For more information see [Vog07],[Vog08]).

Let $(E, \|\cdot\|)$ be a complete separable metric space, χ a compact subset of E and $[\Omega, \Sigma, P]$ a complete probability space. We assume that a deterministic optimization problem

$$(P_0) \min_{x \in \chi} f_0(x)$$

is approximated by a sequence of random problems

$$(P_n) \min_{x \in \chi} f_n(x), n \in N.$$

$f_0 \mid E \to \bar{R}^1$ is a lower semicontinuous function and $f_n \mid E \times \Omega \to \bar{R}^1$ are lower semicontinuous random functions which are supposed to be $(\mathcal{B}(E) \bigotimes \Sigma, \bar{\mathcal{B}}^1)$-measurable. $\mathcal{B}(E)$ denotes the Borel-σ-field of E and $\bar{\mathcal{B}}^1$ the σ-field of Borel sets of $\bar{R}^1$. Finally, we assume that all objective function are (almost surely) proper functions, i.e. functions with values in $(-\infty, +\infty]$ which are not identically ∞.

By Φ_n we denote the optimal value and by Ψ_n the solution set of the random approximate problems (P_n). Correspondingly, by Φ_0 we denote the optimal value and by Ψ_0 the solution set of the deterministic limit problem (P_0). Each solution set can be described as follows:

$$\Psi_0 = \{x \in \chi : f_0(x) - \Phi_0 \le 0\}.$$

$$\Psi_n = \{x \in \chi : f_n(x) - \Phi_n \le 0\}, n \in N_0.$$

We make use of assertions of the following form:

$$\forall \kappa > 0 \ \forall n \in N : P\{\Psi_n \setminus U_{\beta_{n,\kappa}} \Psi_0 \ne \emptyset\} \le \mathcal{H}(\kappa). \tag{2.1}$$

$$\forall \kappa > 0 \ \forall n \in N : P\{\Psi_0 \setminus U_{\beta_{n,\kappa}} \Psi_n \ne \emptyset\} \le \mathcal{H}(\kappa). \tag{2.2}$$

We assume throughout the paper that the sequence $(\beta_{n,\kappa})_{n \in N}$ belongs to the class B of non-increasing sequences of positive numbers and the functions $\mathcal{H}$ belong to the class H of non-increasing functions of the form: $H|R^+ \to R^+$. Of course, one is interested in small confidence sets, hence $(\beta_{n,\kappa})_{n \in N}$ should go to zero as fast as possible and $\mathcal{H}$ should converge to zero as fast as possible if κ tends to infinity. $U_\alpha X$ denotes an open neighborhood of set $X \subset E$ with radius $\alpha : U_\alpha X := \{x \in E : d(x, X) < \alpha\}$.

If the sequence $(\Psi_{n,\kappa})_{n \in N}$ fulfills the relation (1), we call it an *inner approximation in probability* to Ψ_0 with convergence rate $\beta_{n,\kappa}$ and tail behavior function $\mathcal{H}_n$ or just *an inner* $(\beta_{n,\kappa}, \mathcal{H})$*-approximation.* Correspondingly, when a sequence $(\Psi_{n,\kappa})_{n \in N}$ fulfills the relation (2), it is an *outer a approximation in probability* to Ψ_0 with convergence rate $\beta_{n,\kappa}$ and tail behavior function $\mathcal{H}_n$ or in short *an outer* $(\beta_{n,\kappa}, \mathcal{H})$*-approximation.* Since supersets of outer approximations are again outer approximations, one is especially interested in outer approximation which are also inner approximations.

Unfortunately, under reasonable conditions one can only prove inequality (2.1), roughly spoken, that only a subset of the 'true' solution set Ψ_0 is approximated. However, if Ψ_0 is single-valued and the set $\Psi_n, n \in N$, are uniformly bounded, inequality (2.1) implies inequality (2.2). The uniform boundedness condition is satisfied because of the compactness of χ.

Crucial assumptions are uniform concentration-of-measure conditions for the objective functions and conditions about the limit problem, which concern the growth of the objective function.

The growth condition will be described by a function μ which belongs to a set $\Lambda := \{\mu \mid R^+ \to R^+ : \mu$ is increasing, right-continuous, and satisfies $\mu(0) = 0\}$. As it was mentioned above, the constraint set is fixed.

For the reader's convenience we provide below the most important theorems with the short proofs.

Theorem 2.1 (Inner Approximation of the Solution Set) *Assume that the following assumptions are satisfied:*

(2.1) *There exist a function $\mathcal{H} \in H$ and to all $\kappa > 0$ a sequence $(\beta_{n,\kappa})_{n \in N} \in B$ such that*

$$\sup_{n \in N} P\{\sup_{x \in \chi} |f_n(x) - f_0(x)| \ge \beta_{n,\kappa}\} \le \mathcal{H}(\kappa) \tag{2.3}$$

holds.

(2.2) *There exist a function* $\mu \in \Lambda$ *such that for all* $\kappa > 0$

$$\forall x \in \chi \setminus U_\kappa \Psi_0 : f_0(x) \geq \Phi_0 + \mu(\kappa). \tag{2.4}$$

Then for all $\kappa > 0$ *and* $\tilde{\beta}_{n,\kappa} := \mu^{-1}(2\beta_{n,\kappa})$ *the relation* $\sup_{n\in N} P(U_{\tilde{\beta}_{n,\kappa}}\Psi_0 \subset \chi$ and $\Psi_n \setminus U_{\tilde{\beta}_{n,\kappa}}\Psi_0 \neq \emptyset) \leq 2\mathcal{H}(\kappa)$ *holds.*

Proof. Let $\kappa > 0, n \in N$ and $U_{\tilde{\beta}_{n,\kappa}}\Psi_0 \subset \chi$ and $\Psi_n \setminus U_{\tilde{\beta}_{n,\kappa}}\Psi_0 \neq \emptyset$. Then there is $x_n \in \Psi_n$ which does not belong to $U_{\tilde{\beta}_{n,\kappa}}\Psi_0$. Furthermore, choose $x_0 \in \Psi_0$. Because of (2.4) we have $f_0(x_n) - f_0(x_0) \geq \mu(\mu(2\beta_{n,\kappa}))^{-1} \geq 2\beta_{n,\kappa}$. On the other side we have $f_n(x_0) - f_n(x_n) \geq 0$. Consequently, $P\{f_0(x_n) - f_0(x_0) \geq 2\beta_{n,\kappa}\} = P\{-f_n(x_n)+f_0(x_n)+f_n(x_0)-f_0(x_0) \geq 2\beta_{n,\kappa}\} \leq P\{\inf_{x\in\chi\setminus\Psi_0}(f_n(x)-f_0(x)) \leq -\beta_{n,\kappa}\} + P\{\sup_{x\in\Psi_0}(f_n(x) - f_0(x)) \geq \beta_{n,\kappa}\} \leq P\{\inf_{x\in\chi}(f_n(x) - f_0(x)) \leq -\beta_{n,\kappa}\} + P\{\sup_{x\in\chi}(f_n(x) - f_0(x)) \geq \beta_{n,\kappa}\} \leq 2P\{\sup_{x\in\chi}|f_n(x) - f_0(x)| \geq \beta_{n,\kappa}\} \leq 2\mathcal{H}(\kappa)$, and so we have the first condition.□

However, if the solution set is not single-valued, one can obtain only inner approximation and, thus, one can not guarantee that the whole "true" solution set will be covered. Nevertheless, an outer approximation can be constructed as well. To that end we consider "relaxed" problems, where one deals with "relaxed" solution sets. They are accurate only up to a small parameter that depends on n and κ and tends to zero for each κ and $n \to \infty$. With other words, we consider a suitable relaxing sequence $(\rho_{n,\kappa})_{n\in N}$, which tends to zero for each $\kappa > 0$ and consider $\rho_{n,\kappa}$-optimal solutions $\Psi^r_{n,\kappa}$.

$$\Psi^r_{n,\kappa} = \{x \in \chi : f_n(x) - \Phi_n \leq \rho_{n,\kappa}\}, n \in N. \tag{2.5}$$

Theorem 2.2 (Outer Approximation of the Solution Set, relaxation) *Assume that there exist a function* $\mathcal{H} \in \mathrm{H}$ *and to all* $\kappa > 0$ *a sequence* $(\beta_{n,\kappa})_{n\in\mathrm{N}} \in B$ *such that*

$$\sup_{n\in N} P\{\sup_{x\in\chi}|f_n(x) - f_0(x)| \geq \beta_{n,\kappa}\} \leq \mathcal{H}(\kappa) \tag{2.6}$$

holds.

Then for all $\kappa > 0$, $\rho_{n,\kappa} = 2\beta_{n,\kappa}$ $\tilde{\beta}_{n,\kappa} := 2\beta_{n,\kappa}$ *the following relation*

$\sup_{n\geq n_0(\kappa)} P\{U_{\tilde{\beta}_{n,\kappa}}\Psi_0 \subset K$ and $\Psi_0 \setminus (\Psi^r_{n,\kappa} \bigcap U_{\tilde{\beta}_{n,\kappa}}\chi) \neq \emptyset\} \leq 2\mathcal{H}(\kappa)$ *holds.*

Proof. Assume that for given $\kappa > 0, n \in N$ the relation $\Psi_0 \setminus \Psi^r_{n,\kappa} \bigcap U_{\tilde{\beta}_{n,\kappa}}\chi) \neq \emptyset$ is fulfilled. Then there exist $x_{n,\kappa} \in \Psi_0$ which does not belong to $\Psi^r_{n,\kappa} \bigcap U_{\tilde{\beta}_{n,\kappa}}\chi$. Hence,$f_0(x_{n,\kappa}) - \Phi_0 \leq 0$ and $x_{n,\kappa} \in \chi$, but $f_n(x_{n,\kappa}) - \Phi_n > \tilde{\beta}_{n,\kappa} = \rho_{n,\kappa}$.

Because of

$$P\{\inf_{x\in\Psi_0}((f_n(x) - \Phi_n) - (f_0(x) - \Phi_0)) \geq 2\beta_{n,\kappa}\}$$

$$\leq P\{\inf_{x\in\Psi_0}(f_n(x)-f_0(x))\geq\beta_{n,\kappa}\}+P\{-\Phi_n+\Phi_0\geq\beta_{n,\kappa}\}$$

It is not difficult to show that $P\{-\Phi_n+\Phi_0 \geq \beta_{n,\kappa}\}$ is fulfilled, if $P\{\inf_{x\in\mathcal{X}}(-f_n(x)+f_0(x))\geq\beta_{n,\kappa}\}$. Combining it with $P\{\inf_{x\in\Psi_0}(f_n(x)-f_0(x))\geq\beta_{n,\kappa}\}$ we obtain $P\{\sup_{x\in\mathcal{X}}|f_n(x)-f_0(x)|\geq\beta_{n,\kappa}\}$ and, thus, the assumption (2.6) is fulfilled.□

3. Approximation of the regression function. We consider the *regression model with random design* given in (1.2). Assume that (X_i, Y_i) are i.i.d. copies of (X,Y), from which the unknown regression model m will be estimated. Furthermore, it is required that the variable X has a marginal density g. From the definition of the conditional expected value the expression (1.2) can be rewritten as follows:

$$m(x) := \frac{\int yf(x,y)dy}{g(x)}$$

where x is a vector from R and *f(x,y)* is bivariate density. For the sake of convenience we introduce function $r(x) := \int yf(x,y)dy$. Then

$$m(x) := \frac{r(x)}{g(x)}. \tag{3.1}$$

The Nadaraya-Watson estimator has the form

$$\hat{m}(x) = \frac{\hat{r}(x)}{\hat{g}(x)} \tag{3.2}$$

where $\hat{r}$ and $\hat{g}$ are kernel estimates of the functions r and g:

$$\hat{r}(x) := \frac{1}{nh_n}\sum_{i=1}^{n} Y_i K(\frac{x-X_i}{h_n}), \tag{3.3}$$

$$\hat{g}(x) := \frac{1}{nh_n}\sum_{i=1}^{n} K(\frac{x-X_i}{h_n}). \tag{3.4}$$

As it was mentioned in introduction, the Nadaraya-Watson estimator is consistent and asymptotically unbiased. Nevertheless, the quality of the estimation can significantly vary, if we use different kernel functions or different bandwidths. The choice of any of these parameters is a problem of itself. There are many publications considering the methods to make the right choice and it is not our purpose to develop a new one. However, the method we use allows to choose h_n in a way which is optimal in our context. We will discuss this subject briefly in the end of this paper.

Let us consider formulas (3.3) and (3.4) in more details. By K we denote a *kernel function* which is supposed to be a measurable mapping $K: R \to R$ and satisfies the following conditions:

$(K1)$ $\int_{-\infty}^{\infty}|K(x)|dx < \infty$,

$(K2)\int_{-\infty}^{\infty} K(x)dx = 1$,
$(K3)\ K(x) = K(-x)\ \forall x \in R \Rightarrow \int_{-\infty}^{\infty} x^j K(x)dx = 0$ for odd $j \in N$,
$(K4)\ \sup_{x\in R} |K(x)| = C < \infty$.

The multitude of functions satisfing $(K1)$-$(K4)$ is very large. Basically, any density can be used as a kernel. In Appendix we provide a table with the most common kernels. However, different studies show that for large sample sizes the shape of the optimal kernel is unique. For example, in R for L_2 errors, among all positive kernels, the Epanechnikov kernel $K(x) = 0,75(1 - x^2)\mathbb{I}_{\{|x|\le 1\}}$ is the best. Also for the L_1 error, this K is still the best (among all positive kernels) [DL01]. For this reason, in our examples we consider kernel functions with bounded support.

The density g needs special attention. As $|x| \to \infty$, $g(x) \to 0$, the value of m in (3.1) tends to infinity. On the other hand we want the regression function to be finite. So one of the possibilities is to set some bounds on g:

$$\frac{1}{\inf_x g(x)} := C_g < \infty. \tag{3.5}$$

Unfortunately, the density g is unknown, hence the constant C_g can not be given explicitly and should be estimated. We use the Rosenblatt-Parzen estimator $\hat{g}$ as an approximation of g. Applying a result from [Dnb08], we can ensure with the prescribed probability that the approximation error does not exceed a certain value.

$$P(\sup_{x\in\chi} |\hat{g}(x) - g(x)| \ge \beta^g_{n,k}) \le H^g(\kappa). \tag{3.6}$$

Both $\beta^g_{n,k}$ and $H^g(\kappa)$ tend to zero and will be defined later. However, the condition (3.5) can not be fulfilled for any arbitrary x from $\mathbb{R}$. So we choose only such x that satisfy this condition. We assume the the set χ of feasible values of x is defined in such a way that (3.5) is satisfied.

Assume that the regression model was approximated with the Nadaraya-Watson estimator and our main concern is the estimation of the peak of the function. This problem can be described as deterministic approximation problem:

$$(P_0)\ \min_{x\in\chi}(-m(x)) \tag{3.7}$$

where m as in (1.1). It was approximated by the following sequence:

$$(P_n)\ \min_{x\in\chi}(-\hat{m}(x)) \tag{3.8}$$

where $\hat{m}$ as in (3.2).

Firstly, assume that there exists only one maximum, so the solution set is single-valued. Then we can apply Proposition 2.1 in order to construct both the inner and the outer approximation of this set. In other words, we will find $\tilde{\beta}_{n,\kappa}$ and $\mathcal{H}(\kappa)$ for the following inequality:

$$\sup_{n\in N} P(U_{\tilde{\beta}_{n,\kappa}}\Psi_0 \subset \chi \wedge \Psi_n \setminus U_{\tilde{\beta}_{n,\kappa}}\Psi_0 \neq \emptyset) \leq 2\mathcal{H}(\kappa)$$

with Ψ_0 and Ψ_n as a solution set for the problem (3.7) and (3.8) correspondingly.

According to Theorem 2.1 we should show at first that the inequality (2.1) is fulfilled. There are several ways to derive the convergence rate $\beta_{n,\kappa}$ and the tail behavior function $\mathcal{H}(\kappa)$. In this paper we apply the McDiarmid's inequality. This approach gives us quite a good rate of convergence. Unfortunately, it has one important disadvantage: we need the boundedness condition for Y.

$$\sup_y |Y| := C_1 < \infty. \tag{3.9}$$

From the McDiarmid's inequality it is well known that

$$P(|q(X_1,..,X_n) - E[q(X_1,..,X_n)]| \geq \varepsilon) \leq 2e^{-\frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2}},$$

if for all $1 \leq i \leq n$ and a function q the following inequality is satisfied:

$$\sup_{X_1,..,X_i,..,X_n,X_i^* \in A} |q(X_1,..,X_n) - q(X_1,..,X_{i-1},X_i^*,X_{i+1}..,X_n)| \leq c_i.$$

On the other hand, from the the triangle inequality we have

$$\sup_{x\in\chi} |\hat{m}(x) - m(x)| \leq S_1 + S_2 \tag{3.10}$$

with

$$S_1 := |\sup_{x\in\chi} |\hat{m}(x) - m(x)| - E[\sup_{x\in\chi} |\hat{m}(x) - m(x)|]|$$

and

$$S_2 := E[\sup_{x\in\chi} |\hat{m}(x) - m(x)|].$$

The next Lemma gives us the bounds for S_1 and, consequently, the estimation of $H(\kappa)$.

Lemma 3.1 *Assume that conditions (3.5) and (3.9) are satisfied. Furthermore, suppose that*

$$\sup_{x\in\chi} |K(x)| = C_2 < \infty. \tag{3.11}$$

Then for the kernel regression $\hat{m}$ with the bandwidth h_n and kernel K the following inequality holds:

$$P(|\sup_{x\in\chi} |\hat{m}(x) - m(x)| - E[\sup_{x\in\chi} |\hat{m}(x) - m(x)|]| \geq t(\kappa)) \leq H(\kappa) = 2e^{-\frac{t(\kappa)^2 n h_n{}^d}{8C_1^2C_2^2C_g^2}}.$$

Proof. Assume that $q(x, \tilde{X}_1, \tilde{X}_2, ..\tilde{X}_n) := \sup_{x\in\chi} |\hat{m}(x, \tilde{X}_1, \tilde{X}_2, ..\tilde{X}_n) - m(x)|$, where $\tilde{X}_i$ stands for the two-dimensional vector (X_i, Y_i) from the sample. In general q does not depend on $\tilde{X}_1, \tilde{X}_2, ..\tilde{X}_n$. We need them to get an approximation of m, so they have an influence on the form of $\hat{m}$. Assume that there exists another sample and it differs only by a single element $\tilde{X}_i^* := (X_i^*, Y_i^*)$ (instead of $\tilde{X}_i := (X_i, Y_i)$).

Then if the difference

$$d(x) := \sup_{X_1,\dots,X_n \bigwedge X_i^* \in A} |q(x, \tilde{X}_1, ..\tilde{X}_i, ..\tilde{X}_n) - q(x, \tilde{X}_1, ..\tilde{X}_i^*, ..\tilde{X}_n)|$$

$$\leq \sup_{X_1,\dots,X_n \bigwedge X_i^* \in A} \sup_{x\in\chi} |\hat{m}(x, \tilde{X}_1, ..\tilde{X}_i, ..\tilde{X}_n) - \hat{m}(x, \tilde{X}_1, ..\tilde{X}_i^*, ..\tilde{X}_n)|$$

is bounded, we can apply the McDiarmid's inequality.

As $\hat{m}(x) = \frac{\hat{r}(x)}{\hat{g}(x)}$ the difference d(x) can be rewritten as a sum:

$d(x) \leq S_{11} + S_{12}$, with

$$S_{11} := \sup_{X_1,\dots,X_n \bigwedge X_i^* \in A} \sup_{x\in\chi} \left|\frac{\hat{r}(x, \tilde{X}_1, ..\tilde{X}_i, ..\tilde{X}_n) - \hat{r}(x, \tilde{X}_1, ..\tilde{X}_i^*, ..\tilde{X}_n)}{\hat{g}(x, \tilde{X}_1, ..\tilde{X}_i, ..\tilde{X}_n)}\right|$$

$$S_{12} := \sup_{X_1,\dots,X_n \bigwedge X_i^* \in A} \sup_{x\in\chi} \left|\frac{\hat{r}(x, \tilde{X}_1, ..\tilde{X}_i^*, ..\tilde{X}_n)}{\hat{g}(x, \tilde{X}_1, ..\tilde{X}_i^*, ..\tilde{X}_n)} \cdot \frac{\hat{g}(x, \tilde{X}_1, ..\tilde{X}_i^*, ..\tilde{X}_n) - \hat{g}(x, \tilde{X}_1, ..\tilde{X}_i, ..\tilde{X}_n)}{\hat{g}(x, \tilde{X}_1, ..\tilde{X}_i, ..\tilde{X}_n)}\right|.$$

At first we consider S_{11}. Making use of assumptions (3.5), (3.9) and (3.11) we have

$$S_{11} := \sup_{X_1,\dots,X_n \bigwedge X_i^* \in A} \sup_{x\in\chi} \left|\frac{\hat{r}(x, \tilde{X}_1, ..\tilde{X}_i, ..\tilde{X}_n) - \hat{r}(x, \tilde{X}_1, ..\tilde{X}_i^*, ..\tilde{X}_n)}{\hat{g}(x, \tilde{X}_1, ..\tilde{X}_i, ..\tilde{X}_n)}\right|$$

$$\leq \sup_{X_1,\dots,X_n \bigwedge X_i^* \in A} \sup_{x\in\chi} C_g |\hat{r}(x, \tilde{X}_1, ..\tilde{X}_i, ..\tilde{X}_n) - \hat{r}(x, \tilde{X}_1, ..\tilde{X}_i^*, ..\tilde{X}_n)| =$$

$$\sup_{X_1,\dots,X_n \bigwedge X_i^* \in A} \sup_{x\in\chi} \frac{C_g}{nh_n} |Y_i K(\frac{x - X_i}{h_n}) - Y_i^* K(\frac{x - X_i^*}{h_n})| \leq \frac{2}{nh_n} C_g C_1 C_2.$$

Further, we estimate the summand S_{12}. The first factor of it is

$$\sup_{X_1,\dots,X_n \bigwedge X_i^* \in A} \sup_{x\in\chi} \left|\frac{\hat{r}(x, \tilde{X}_1, ..\tilde{X}_i^*, ..\tilde{X}_n)}{\hat{g}(x, \tilde{X}_1, ..\tilde{X}_i^*, ..\tilde{X}_n)}\right| = \sup_{x_1,\dots,x_n \bigwedge x_i' \in A} \sup_{x\in\chi} \left|\frac{\sum_{i=1}^n Y_i K(\frac{x - X_i}{h_n})}{\sum_{i=1}^n K(\frac{x - X_i}{h_n})}\right| \leq C_1.$$

The second factor is similar to S_{11}:

$$\sup_{X_1,\dots,X_n \bigwedge X_i^* \in A} \sup_{x\in\chi} \left|\frac{\hat{g}(x, \tilde{X}_1, ..\tilde{X}_i^*, ..\tilde{X}_n) - \hat{g}(x)}{\hat{g}(x)}\right| \leq \frac{2C_2 C_g}{nh_n}.$$

Thus, the difference d(x) is bounded by $\frac{4C_1C_2C_g}{nh_n}$ and because of McDiarmid's inequality the conclusion is obtained. $\square$

The form of the function $H(\kappa)$ is known. Now we are going to find an estimation for S_2. In order to achieve it, we expand S_2:

$$S_2 := E[\sup_{x \in \chi} |\hat{m}(x) - m(x)|] \leq S_{21} + S_{22} + S_{23} + S_{24}$$

with

$$S_{21} := E[\sup_{x \in \chi} |\frac{\hat{r}(x) - E[\hat{r}(x)]}{\hat{g}(x)}|],$$

$$S_{22} := E[\sup_{x \in \chi} |\frac{E[\hat{r}(x)]}{E[\hat{g}(x)]} \frac{E[\hat{g}(x)] - \hat{g}(x)}{\hat{g}(x)}|],$$

$$S_{23} := \sup_{x \in \chi} |\frac{E[\hat{r}(x)] - r(x)}{\hat{g}(x)}|,$$

$$S_{24} := \sup_{x \in \chi} |\frac{E[\hat{r}(x)]}{E[\hat{g}(x)]} \frac{g(x) - E[\hat{g}(x)]}{\hat{g}(x)}|.$$

For the further calculations we employ a Fourier transform to the kernel (as in [Par62]): $k(u) := \int_{-\infty}^{\infty} e^{\imath u y} K(y) dy, \forall u \in \mathbb{R}$. In Appendix we provide a table with the formulas of k for the most common kernels. Because of the complete integrability of kernel $(K1)$, the reverse transformation is also possible: $K(u) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\imath u y} k(y) dy, \forall u \in \mathbb{R}$. Furthermore, we consider the Fourier transformation for the approximated functions $\hat{r}$ and $\hat{g}$:

$$\hat{r}(x) := \frac{1}{nh_n} \sum_{l=1}^{n} (\frac{1}{2\pi} \int_{-\infty}^{\infty} Y_l e^{-\imath \frac{(x-X_l)}{h_n} u} k(u)) du = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\imath x u} k(h_n u) (\frac{1}{n} \sum_{l=1}^{n} Y_l e^{\imath X_l u}) du.$$

We denote by β the following sum: $\beta(u) := \frac{1}{n} \sum_{l=1}^{n} Y_l e^{\imath X_l u}$. Then

$$\hat{r}(x) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\imath x u} k(h_n u) \beta(u) du.$$

In the same way we obtain the result for $\hat{g}$:

$$\hat{g}(x) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\imath x u} k(h_n u) \alpha(u) du,$$

where $\alpha(u) := \frac{1}{n} \sum_{l=1}^{n} e^{\imath X_l u} \forall u \in \mathbb{R}$.

Lemma 3.2 *Assume that conditions (3.5), (3.9) and (3.11) are satisfied. Furthermore, suppose that*

$$\int_{-\infty}^{\infty} |k(u)| du =: C_3 < \infty \tag{3.12}$$

$$\mathbf{var}(|Y|) =: (C_4)^2 < \infty \tag{3.13}$$

Then the following inequality holds:

$$S_{21} = E[\sup_{x\in\chi}|\frac{\hat{r}(x) - E[\hat{r}(x)]}{\hat{g}(x)}|] \leq \frac{C_g C_1 C_3}{2\pi\sqrt{n}h_n}.$$

Proof . Firstly we apply condition (3.5):

$$S_{21} \leq C_g E[\sup_{x\in\chi}|\hat{r}(x) - E[\hat{r}(x)]|].$$

Employing Jensen's inequality and the Fourier transformation defined above we obtain:

$$E^2[\sup_{x\in\chi}|\hat{r}(x) - E[\hat{r}(x)]|] \leq E[\sup_{x\in\chi}|\hat{r}(x) - E[\hat{r}(x)]|^2] =$$

$$= E[\sup_{x\in\chi}|\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\imath xu}k(h_n u)\beta(u)du - E[\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\imath xu}k(h_n u)\beta(u)du]|^2]$$

As $|e^{-\imath xu}| = 1$ the expression will be reduced to

$$E[\sup_{x\in\chi}|\hat{r}(x) - E[\hat{r}(x)]|^2] \leq E[|\frac{1}{2\pi}\int_{-\infty}^{\infty}|k(h_n u)|(\beta(u) - E[\beta(u)])du|^2]$$

With the theorem of Fubini, Cauchy-Schwarz inequality and integrability of $|k(h_n u)|(\beta(u) - E[\beta(u)])du|^2$ we have

$$E^{\frac{1}{2}}[|\frac{1}{2\pi}\int_{-\infty}^{\infty}|k(h_n u)|(\beta(u) - E[\beta(u)])du|^2] \leq \frac{1}{2\pi}\int_{-\infty}^{\infty}E^{\frac{1}{2}}[|k(h_n u)|^2(\beta(u) - E[\beta(u)])^2du] =$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}|k(h_n u)|\sqrt{E|\beta(u) - E[\beta(u)]|^2}du = \frac{1}{2\pi}\int_{-\infty}^{\infty}|k(h_n u)|\sqrt{\mathbf{var}(\beta(u))}du.$$

Furthermore,

$$\mathbf{var}(\beta(u)) = \frac{1}{n^2}\sum_{k=1}^{n}\mathbf{var}(Y_k e^{\imath u X_k}) = \frac{1}{n}\mathbf{var}(|Y e^{\imath u X}|) \leq \frac{1}{n}\mathbf{var}(|Y|) \leq \frac{(C_4)^2}{n}.$$

Applying condition (3.12) we obtain the conclusion of this Lemma:

$$S_{21}^{(1)} \leq C_g E[\sup_{x\in\chi}|\hat{r}(x) - E[\hat{r}(x)]|] \leq \frac{C_g}{2\pi}\int_{-\infty}^{\infty}|k(h_n u)|\sqrt{\mathbf{var}(\beta(u))}du \leq$$

$$\leq \frac{C_g C_1}{2\pi\sqrt{n}h_n}\int_{-\infty}^{\infty}|k(y)|dy \leq \frac{C_g C_3 C_4}{2\pi\sqrt{n}h_n}.\square$$

In the same way we proceed with S_{22}.

Lemma 3.3 *Assume that conditions (3.5), (3.6), (3.11)and (3.12) are satisfied. Furthermore, suppose that*

$$E[|Y|] := C_5 < \infty. \tag{3.14}$$

Then the following inequality holds:

$$S_{22} = E[\sup_{x\in\chi} |\frac{E[\hat{r}(x)]}{E[\hat{g}(x)]} \frac{E[\hat{g}(x)] - \hat{g}(x)}{\hat{g}(x)}|] \leq \frac{C_g C_3 C_5}{2\pi\sqrt{n}h_n}.$$

Proof. This expression is an expected value of a product. Moreover:

$$E[\sup_{x\in\chi} |\frac{E[\hat{r}(x)]}{E[\hat{g}(x)]} \frac{E[\hat{g}(x)] - \hat{g}(x)}{\hat{g}(x)}|] \leq E[\sup_{x\in\chi} |\frac{E[\hat{r}(x)]}{E[\hat{g}(x)]}| \times \sup_{x\in\chi} |\frac{E[\hat{g}(x)] - \hat{g}(x)}{\hat{g}(x)}|].$$

Let us consider the first multiplier. It can be estimated as follows.

$$\sup_{x\in\chi} |\frac{E[\hat{r}(x)]}{E[\hat{g}(x)]}| \leq \sup_{x\in\chi} |\frac{\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\imath xu} k(h_n u) E[\beta(u)]}{\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\imath xu} k(h_n u) E[\alpha(u)]}| \leq \sup_{x\in\chi} |\frac{E[\beta(u)]}{E[\alpha(u)]}| \leq C_5.$$

The rest of the proof is similar to Lemma 3.2. □

The assumption we used in the previous two Lemmas involves the absolute integrability of k. It is easy to see that this condition is true almost for any k. In the Appendix we provide estimations of C_3 for the most common kernels.

For the next Lemma we need r and g to be continuous and two-times differentiable. Moreover, we assume the derivatives to be bounded. The main difficulty is that both r and g are unknown. There are several ways to estimate these derivatives. The easiest way is to employ for the following reasons the *kernel derivative estimators*. They can be defined by differentiating the kernel function sequence with respect to x. If the approximation is sufficiently smooth and the bandwidth sequence is correctly tuned then these estimators will converge to the corresponding derivatives of r and g [Hdl89], [Zgl02].

$$\hat{r}''(x) := \frac{1}{nh_n^3}\sum_{i=1}^{n} Y_i K''(\frac{x - X_i}{h_n}) \text{ and } \hat{g}''(x) := \frac{1}{nh_n^3}\sum_{i=1}^{n} K''(\frac{x - X_i}{h_n})$$

Unfortunately, application of this estimation with $h_n \to 0$ makes $\beta_{n,\kappa}$ tending to infinity (see Lemma 3.6). As this problem needs some additional study, we will consider it elsewhere. However, it worth to mention that estimation of density's derivative appeared in many papers (see [Jnl92]). Besides, g'' is easier to obtain as we need no special knowledge about Y.

Lemma 3.4 *Assume that the condition (3.5) is fulfilled. Furthermore, suppose that the functions r and g and the kernel function satisfy the following assumptions:*

$$r(x), g(x) \in C^2(R), \tag{3.15}$$

$$\int_{-\infty}^{\infty} x^2 K(x)dx = C_6 < \infty, \tag{3.16}$$

$$\exists r'' \textit{ and } |r''| = C_7 < \infty. \tag{3.17}$$

Then the following inequality holds:

$$S_{23} = \sup_{x\in\chi} |\frac{E[\hat{r}(x)] - r(x)}{\hat{g}(x)}| \leq \frac{h_n C_g C_6 C_7}{2}.$$

Proof. At first we rewrite the expected value and apply assumption (3.5):

$$S_{23} \leq C_g \sup_{x\in\chi} |E[\hat{r}(x)] - r(x)| = \frac{C_g}{h_n} \sup_{x\in\chi} |\int_{-\infty}^{\infty} K(\frac{x-y}{h_n}) r(y)dy - r(x)|$$

Since $K(u) = K(-u)$, we can use the following substitution: $\frac{x-y}{h_n} = -u$ and obtain

$$S_{23} \leq \frac{C_g}{h_n} \sup_{x\in\chi} |\int_{-\infty}^{\infty} K(u) r(x + uh_n)du - r(x)|.$$

We use a second order Taylor expansion of r around x with the remainder of Lagrange form ($\xi_{x,u} \in [x, x+u]$).

$$r(x + uh_n) = r(x) + (uh_n)r'(x) + \frac{1}{2}(uh_n)^2 r''(\xi).$$

Further,

$$S_{23} \leq \frac{C_g}{h_n} \sup_{x\in\chi} |\int_{-\infty}^{\infty} K(u)(r(x+uh_n) - r(x))du| =$$

$$= \frac{C_g}{h_n} \sup_{x\in\chi} |\int_{-\infty}^{\infty} K(u)(r(x) + (uh_n)r'(x) + \frac{1}{2}(uh_n)^2 r''(\xi) - r(x))du| =$$

$$= \frac{C_g}{h_n} \sup_{x\in\chi} |\int_{-\infty}^{\infty} K(u)(uh_n)r'(x) + \frac{1}{2}K(u)(uh_n)^2 r''(\xi)udu| \leq$$

$$\leq \frac{C_g}{h_n} \sup_{x\in\chi} |\int_{-\infty}^{\infty} \frac{1}{2}K(u)(uh_n)^2 r''(\xi)du| \leq$$

$$\frac{h_n C_g \mathcal{H}_r(\xi)}{2} |\sup_{x\in\bar{\chi}} |\int_{-\infty}^{\infty} u^2 K(u)du| = \frac{h_n C_g C_6 C_7}{2}.$$ □

Lemma 3.5 *Assume that conditions (3.5), (3.15) and (3.16) are fulfilled. Besides, suppose that*

$$\exists g'' \textit{ and } |g''| = C_8 < \infty. \tag{3.18}$$

Then the following inequality holds:

$$S_{24} := \sup_{x\in\chi} | \frac{E[\hat{r}(x)]}{E[\hat{g}(x)]} \frac{g(x) - E[\hat{g}(x)]}{\hat{g}(x)} | \leq \frac{h_n C_g C_5 C_6 C_8}{2}.$$

Proof. This follows by the same method as in Lemma 3.3 and 3.4. □

Note that the Taylor expansion can be continued, if all the further derivatives of r and g exist and are bounded. In view of condition $(K3)$ only even derivatives are needed. Besides, we have to estimate the following integral:

$$\int_{-\infty}^{\infty} u^{2l} K(u) du, \; for \; l \geq 2.$$

It worth to mention that for the most common kernels these values constitute a decreasing sequence tending to zero.

Now we have all the knowledge we need about the parameters in (2.3).

Lemma 3.6 *Assume that the assumptions (3.5),(3.9) and (3.11)-(3.18) are satisfied. Then for the kernel regression $\hat{m}$ with the bandwidth h_n and kernel K the following inequality holds:*

$$P(\sup_{x\in\chi} |\hat{m}(x) - m(x)| \geq \beta_{n,\kappa}) \leq \mathcal{H}(\kappa)$$

where

$$\beta_{n,\kappa} = \frac{\kappa}{\sqrt{n} h_n} + \frac{C_g C_1 C_3}{\pi \sqrt{n} h_n} + \frac{h_n C_g C_6 C_7}{2} + \frac{h_n C_g C_5 C_6 C_8}{2}$$

and

$$\mathcal{H}(\kappa) = 2e^{-\frac{\kappa^2}{8C_1^2 C_2^2 C_g^2}}.$$

Proof. We apply the triangle inequality as in (3.12):

$$\sup_{n\in N} P(\sup_{x\in\chi} |\hat{m}(x) - m(x)| \geq \beta_{n,\kappa}) \leq \sup_{n\in N} P(S_1 + S_2 \geq \beta_{n,\kappa}) \leq$$

$$\sup_{n\in N} P(S_1 + \frac{C_g C_1 C_3}{\pi\sqrt{n} h_n} + \frac{h_n C_g C_6 C_7}{2} + \frac{h_n C_g C_1 C_6 C_8}{2} \geq \beta_{n,\kappa}) =$$

$$\sup_{n\in N} P(| \sup_{x\in\chi} |\hat{m}(x) - m(x)| - E[\sup_{x\in\chi} |\hat{m}(x) - m(x)|]| \geq \frac{\kappa}{\sqrt{n} h_n}) \leq 2e^{-\frac{\kappa^2}{8C_1^2 C_2^2 C_g^2}}.$$

Making use of the result of Lemma 3.1 with $t(\kappa) = \frac{\kappa}{\sqrt{n} h_n}$, the proof is completed. □

Finally, we can construct a universal confidence set for the mode of the regression model:

Theorem 3.7 *Assume that the conditions (3.5),(3.9) and (3.11)-(3.18) are satisfied. Furthermore, assume that there exist a function $\mu \in \Lambda$ such that*

$$\kappa > 0 \ \forall x \in \chi \setminus U_\kappa \Psi_0 : -m(x) \geq \Phi_0 + \mu(\kappa).$$

Then the confidence set Ψ_n *for the solution set of problem (3.7)* Ψ_0 *can be derived as follows*:

$$\sup_{n \in N} P(U_{\tilde{\beta}_{n,\kappa}} \Psi_0 \subset K \wedge \Psi_n \setminus U_{\tilde{\beta}_{n,\kappa}} \Psi_0 \neq \emptyset) \leq 4e^{-\frac{\kappa^2}{8C_1^2 C_2^2 C_g^2}}$$

and

$$\tilde{\beta}_{n,\kappa} = \mu^{-1}(2(\frac{\kappa}{\sqrt{n}h_n} + \frac{C_g C_1 C_3}{\pi\sqrt{n}h_n} + \frac{h_n C_g C_6 C_7}{2} + \frac{h_n C_g C_1 C_6 C_8}{2})).$$

Proof. This theorem follows from Theorem 2.1 and Lemma 3.6.□

In the end we have to go back to the condition (3.6) and give the estimation for the parameters in there.

Lemma 3.8 *Assume that the assumptions (3.11), (3.12), (3.15),(3.16) and (3.18) are satisfied. Then for the kernel density* $\hat{g}$ *with the bandwidth* h_n *and kernel* K *the following inequality holds:*

$$P(\sup_{x \in \chi} |\hat{g}(x) - g(x)| \geq \beta^g_{n,k}) \leq 2e^{\frac{-\kappa^2}{2C_2^2}}$$

with

$$\beta^g_{n,k} = \frac{\kappa}{\sqrt{n}h_n} + \frac{C_3}{2\pi\sqrt{n}h_n} + \frac{h_n C_g C_1 C_6 C_8}{2}.$$

Proof. This follows by the same method as in Lemma 3.6.□

4. Approximation of the multimodal regression function. In the previous section we dealt with problem (3.6) and obtained a confidence set for its solution. Unfortunately, the assumption of the uniqueness of the mode is usually too restrictive. So we have to allow the regression function to be multimodal, i.e. to have several peaks. However, it leads immediately to the problem that only an inner approximation of the solution set can be achieved directly as it was done in Section 3. In the general case an outer approximation is not possible and we have to deal with the "relaxed" problems, or in our case with "relaxed" ($\rho_{n,\kappa}$-optimal) solution $\Psi^r_{n,\kappa}$ as in (2.5).

The Theorem 2.2 provides a criterion that makes it offers the possibility to construct a universal confidence set in this case. It is obvious that the condition (2.6) is equivalent to the condition (2.3). Furthermore, the Theorem 3.7 yields all the information about it. Thus, we are able to obtain an outer approximation without any further calculations.

Theorem 4.1 *Assume that the conditions (3.5),(3.9) and (3.11)-(3.18) are satisfied. Then for all* $\kappa > 0$ *and* $\rho_{n,\kappa} = \tilde{\beta}_{n,\kappa} = 2\beta_{n,\kappa}$ *the solution set of the problem (P)* Ψ_0 *can be approximated by* $\Psi^r_{n,\kappa}$ *so that* $\Psi^r_{n,\kappa}$ *covers* Ψ_0 *with the probability* $4e^{-\frac{\kappa^2}{8C_1^2C_2^2C_g^2}}$:

$\sup_{n\geq n_0(\kappa)} P\{U_{\tilde{\beta}_{n,\kappa}}\Psi_0 \subset \chi \text{ and } \Psi_0 \setminus (\Psi^r_{n,\kappa} \bigcap U_{\tilde{\beta}_{n,\kappa}}\chi) \neq \emptyset\} \leq 4e^{-\frac{\kappa^2}{8C_1^2C_2^2C_g^2}}$

and $\beta_{n,\kappa} = \mu^{-1}(2(\frac{\kappa}{\sqrt{n}h_n} + \frac{C_gC_1C_3}{\pi\sqrt{n}h_n} + \frac{h_nC_gC_6C_7}{2} + \frac{h_nC_gC_1C_6C_8}{2}))$.

Choice of the bandwidth. The quality of estimation strongly depends on the choice of the kernel and its bandwidth. In Section 3 we have already discussed the optimal selection of the K. However, according to Silverman (1986) the choice of h_n is much more important for the behavior of $\hat{g}$ and $\hat{m}$ then the choice of kernel. Small values of h_n make the estimate look "wiggly" and show spurious features, whereas big values of h_n will lead to an estimate which is too smooth in the sense that it is too biased and may not reveal structural features. Due to these reasons a lot of research was done to find objective, data-driven bandwidth selection methods. There exist several measures which can be used to access the goodness of the estimation. This variety naturally lead to different definitions which h_n is optimal.

Meanwhile, if we consider the convergence rate $\tilde{\beta}_{n,\kappa}$ (see Lemma 3.6), we will see that it also depends on the bandwidth h_n:

$$\beta_{n,\kappa} = \frac{\kappa}{\sqrt{n}h_n} + \frac{C_gC_1C_3}{\pi\sqrt{n}h_n} + \frac{h_nC_gC_6C_7}{2} + \frac{h_nC_gC_5C_6C_8}{2}.$$

Considering this formula summandwise, we see that for the first two fractions

$$\frac{\kappa}{\sqrt{n}h_n} \text{ and } \frac{C_gC_1C_3}{\pi\sqrt{n}h_n}$$

we need h_n to remain large enough. For the rest of the formula

$$\frac{h_nC_gC_6C_7}{2} \text{ and } \frac{h_nC_gC_5C_6C_8}{2}$$

h_n should tend to zero. This contradiction can be used as an additional information by choosing the optimal value of h_n. Moreover, the bandwidth that is optimal in the sense of this method can be determined by "balancing" this two parts:

$$\frac{\kappa}{\sqrt{n}h_n} + \frac{C_gC_1C_3}{\pi\sqrt{n}h_n} = \frac{h_nC_gC_6C_7}{2} + \frac{h_nC_gC_5C_6C_8}{2}.$$

It is not very difficult to see that h_n should be

$$h_n^{opt} = \sqrt{\frac{2\pi\kappa + 2C_gC_1C_3}{\pi\sqrt{n}C_gC_6(C_7 + C_5C_8)}}.$$

Appendix. Let us say a few words about the constants from this paper. Some of them depend on the sample, so we can compute or estimate them directly, as, for example, C_1, C_4 or C_5.

However, it is of interest to consider in more detail those constants that depend mostly on the form of the function K. We summarize the results for the most common kernels in the table. It is worth pointing out that almost all the functions K we consider have bounded support (the advantages of such a form were discussed in Section 3). The only exception is the last kernel - the Gaussian kernel. We consider it only to show that all the constants can be easily determined for any arbitrary K. The most interesting is the third line of the table, as we provide there the results for the Epanechnikov kernel and, as it was already mentioned, this form of K usually leads to the best approximation.

Kernel	**K(u)**	$\mathbf{C_2}$	**k(u)**	$\mathbf{C_3}$	$\mathbf{C_6}$
Uniform	$\frac{1}{2}(\mathbb{I}_{(\lvert u\rvert\le 1)})$	$\frac{1}{2}$	$\frac{\sin(u)}{u}$	$n\backslash a$	$\frac{1}{3}$
Triangle	$(1-\lvert u\rvert)\mathbb{I}_{(\lvert u\rvert\le 1)}$	1	$(\frac{\sin(\frac{u}{2})}{\frac{u}{2}})^2$	2π	$\frac{1}{6}$
Epanechnikov	$\frac{3}{4}(1-u)^2(\mathbb{I}_{(\lvert u\rvert\le 1)})$	$\frac{3}{4}$	$\frac{3(\sin(u)-u\cos(u))}{u^3}$	$\frac{3\pi}{2}$	$\frac{1}{5}$
Quadratic	$\frac{15}{16}(1-u^2)^2(\mathbb{I}_{(\lvert u\rvert\le 1)})$	$\frac{15}{16}$	$\frac{-15(\sin(u)(u^2-\ 3)+3u\cos(u))}{u^5}$	$-\frac{15\pi}{8}$	$\frac{1}{7}$
Triweight	$\frac{35}{32}(1-u^2)^3(\mathbb{I}_{(\lvert u\rvert\le 1)})$	$\frac{35}{32}$	$\frac{105(\cos(u)(u^3-15u)+\sin(u)(15-6u^2))}{u^7}$	$\frac{35\pi}{16}$	$\frac{1}{9}$
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2)$	$\frac{1}{\sqrt{2\pi}}$	$\exp(-\frac{1}{2}u^2)$	$\sqrt{2\pi}$	1

References:

[DL01] Devroye, L. and G. Lugosi: Combinatorial Methods in Density Estimation. Springer, 2001.

[Dnb08] Dünnbier, Anne, Diplom: Konfidenzintervalle für Modalwerte, 2008. TU Ilmenau.

[HW91] Härdle, Wolfgang : Applied nonparametric regression. Cambridge Univ. Press, 1991.

[Jn92] Jones, M.C. : Differences and Derivatives in Kernel Estimation. Metrika (1992) 39:335-340.

[MP07] Massart, Pascal: Concentration Inequalities and Model Selection. Ecole d'Eté de Probabilités de Saint-Flour XXXIII-2003. Springer, 2007

[Par62] Parzen, E.: On estimation of a probability density function and mode, Ann. Math. Stat. 33, pp. 1065-1076, 1962.

[Pf04] Pflug, Georg Ch.: Stochastic Optimization and Statistical Inference. In: Stochastic Programming, Handbooks in Operation Research and Management Science, pp. 427-482, 2003.

[Trl93] Turlach, Berwin A.: Bandwidth Selection in Kernel Density Estimation. Discussion Paper, 1993.

[Si86] Silverman, B.W.: Density estimation for Statistics and Data Analysis. Chapman and Hall, 1986.

[Str94] Stroock, D.W.: A Concise Intoduction to the theory of Integration. Birkhäuser, 1994.

[Vog07] Vogel, Silvia: Universal Confidence Sets for Solutions of Optimization Problems. SIAM Journal on Optimization 19 (2008), 3, 1467 - 1488.

[Vog08] Vogel, Silvia: Confidence Sets and Convergence of Random Functions. In: Festschrift in Celebration of Prof. Dr. Wilfried Grecksch's 60th Birthday; C Tammer, F. Heyde (Hrsg.), Shaker-Verlag 2008

[Zi01] Ziegler, Klaus: On approximations to the bias of the Nadaraya-Watson regression estimator. Nonparametric Statistic, Vol. 13, pp. 583-589.

[Zi02] Ziegler, Klaus: On nonparametric kernel estimation of the mode of the regression function in the random design model. Journal of Nonparametric Statistics, 14: 749-774,2002.